Published online 22 October 2013 *Nucleic Acids Research, 2014, Vol. 42, No. 2* **1341–1353**
doi:10.1093/nar/gkt922

Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases

Ole Niewoehner[1], Martin Jinek[1,2] and Jennifer A. Doudna[1,2,3,4,*]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA, [2]Howard Hughes Medical Institute, University of California, Berkeley, California 94720, USA, [3]Department of Chemistry, University of California, Berkeley, California 94720, USA and [4]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

Received June 23, 2013; Revised September 17, 2013; Accepted September 20, 2013

ABSTRACT

In many bacteria and archaea, small RNAs derived from clustered regularly interspaced short palindromic repeats (CRISPRs) associate with CRISPR-associated (Cas) proteins to target foreign DNA for destruction. In Type I and III CRISPR/Cas systems, the Cas6 family of endoribonucleases generates functional CRISPR-derived RNAs by site-specific cleavage of repeat sequences in precursor transcripts. CRISPR repeats differ widely in both sequence and structure, with varying propensity to form hairpin folds immediately preceding the cleavage site. To investigate the evolution of distinct mechanisms for the recognition of diverse CRISPR repeats by Cas6 enzymes, we determined crystal structures of two *Thermus thermophilus* Cas6 enzymes both alone and bound to substrate and product RNAs. These structures show how the scaffold common to all Cas6 endonucleases has evolved two binding sites with distinct modes of RNA recognition: one specific for a hairpin fold and the other for a single-stranded 5′-terminal segment preceding the hairpin. These findings explain how divergent Cas6 enzymes have emerged to mediate highly selective pre-CRISPR-derived RNA processing across diverse CRISPR systems.

INTRODUCTION

Clustered regularly interspaced short palindromic repeats–CRISPR-associated (CRISPR–Cas) systems are bacterial adaptive immune systems that use CRISPR-derived RNAs (crRNAs) together with Cas proteins to defend against invasive genetic elements including bacteriophages or plasmids (1–4). Found in many bacterial and most archaeal genomes, CRISPR loci are transcribed as long pre-crRNAs that are processed enzymatically into ~60-nt mature crRNAs (5). In association with Cas proteins, crRNAs target foreign genetic elements for destruction by base pairing to complementary sequences in phage or plasmid DNA.

Ribonucleases belonging to the Cas6 clade of Repeat-Associated Mysterious Proteins (RAMP), found within Type I and III CRISPR–Cas systems, share the ability to recognize and cleave a single phosphodiester bond in a short repeated sequence of the pre-crRNA transcript (1–4,6). Cas6-mediated cleavage produces mature crRNAs bearing a unique spacer-derived guide sequence flanked by repeat-derived sequences on the 5′ and 3′ ends (5,7,8). Cas6 enzymes are metal-independent nucleases that catalyze RNA cleavage via a mechanism involving a 2′–3′ cyclic intermediate (8,9). Structural studies have shown that Cas6 enzymes share a common ferredoxin or RNA recognition motif (RRM) fold despite having widely divergent amino acid sequences (7,8,10–12). This sequence divergence has been thought to be responsible for the ability of Cas6 enzymes to recognize different kinds of RNA substrates. Many Type I CRISPR repeat sequences have the potential to form stable hairpin structures (13), which create the major-groove binding sites for *Pseudomonas aeruginosa* Cas6f (PaCas6f, also known as Csy4) and *Thermus thermophilus* Cas6e (TtCas6e, also known as Cse3 or CasE) enzymes (8,10,11,14). By contrast, a subset of Type I and Type III CRISPR systems derive their crRNAs from loci in which the repeat sequences are predicted to be unstructured. Crystallographic studies of *Pyrococcus furiosus* Cas6 (PfCas6), a prototypical Cas6 enzyme that cleaves an unstructured repeat sequence, have revealed that the ribonuclease recognizes a 5′ terminal region of the repeat at a considerable distance upstream of the cleavage site (15).

*To whom correspondence should be addressed. Tel: +1 510 643 0225; Fax: +1 510 643 0080; Email: doudna@berkeley.edu
Present address:
Ole Niewoehner and Martin Jinek, Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland.

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

To determine how the Cas6 enzyme family has evolved distinct RNA recognition capabilities based on a conserved structural core, we investigated two Cas6 enzymes associated with CRISPR loci in which the crRNA repeat sequences are predicted to form weak hairpin structures. These *T. thermophilus* enzymes, hereafter referred to as TtCas6A and TtCas6B, are each predicted to recognize a four-base pair stem-loop just upstream of the cleavage site within pre-crRNA transcripts. Five crystal structures of TtCas6A and TtCas6B, both alone and in complex with their cognate substrate and product RNAs, show that although TtCas6A and TtCas6B share nearly identical structures, they use distinct modes of RNA recognition. Furthermore, binding studies and kinetic assays, together with comparisons with related Cas6 crystal structures, reveal a binding mechanism in which both the stem-loop of the repeat RNA and a single-stranded upstream 5′ segment are indispensable for substrate recognition, implying a functional link between two distinct RNA binding surfaces in Cas6 enzymes. These findings provide an explanation for the evolutionary relationship between Cas6 enzymes with orthogonal substrate recognition capabilities and suggest mechanisms by which distinct substrate binding modes can evolve from a single protein scaffold.

MATERIALS AND METHODS

Protein expression and purification

The genes encoding TtCas6A (TTHA0078) and TtCas6B (TTHB231) were amplified from genomic DNA of *T. thermophilus* HB8 and cloned into customized pET-based expression vectors (pEC-K-His and pEC-K-His-MBP) using ligation-independent cloning, resulting in protein constructs in which TtCas6A or TtCas6B were fused downstream of a hexahistidine affinity tag (pEC-K-His) or a hexahistidine-maltose-binding protein (MBP) tag (pEC-K-His-MBP) and a tobacco etch virus protease cleavage site. R22A, R129A and H37A mutants of TtCas6A and the H23A and H42A mutants of TtCas6B were generated using the QuikChange site-directed mutagenesis method (Agilent), and point mutations were verified by DNA sequencing. Expression plasmids were transformed into *Escherichia coli* BL21 Rosetta 2 (DE3) cells (Novagen), and protein expression was induced using 200 μM IPTG at an optical cell density (OD_{600}) of ~0.7, followed by shaking at 18°C for 16 h. Cells were harvested and lysed by sonication in 20 mM Tris-HCl (pH 8.0), 250 mM KCl, 20 mM imidazole, supplemented with 0.2 mg/ml lysozyme and protease inhibitors (Roche). For cleavage assays and crystallographic purposes, the proteins were purified as N-terminal hexahistidine fusions as follows. The cleared lysate was incubated with Ni-NTA affinity resin (Qiagen) in 20 mM Tris-HCl (pH 8.0), 250 mM KCl and 20 mM imidazole, and hexahistidine-tagged protein was eluted with 250 mM imidazole. Eluted proteins were then dialyzed against 20 mM Tris-HCl (pH 7.5), 150 mM KCl and 1 mM TCEP in presence of tobacco etch virus protease, which was followed by a second Ni-NTA step to remove the hexahistidine tag. Proteins were further purified by size-exclusion chromatography using a Superdex 75 (16/60) column (GE Life Sciences) in 20 mM Tris-HCl (pH 7.0) and 250 mM KCl. For electrophoretic mobility shift assays (EMSAs), the proteins were expressed as hexahistidine–MBP fusions and purified without prior proteolytic cleavage and removal of the MBP tag by size-exclusion chromatography using a Superdex 200 (16/60) column in the same buffer.

RNA transcription and labeling

The non-cleavable substrate mimicking R1 stem-loop (Δ1–14, Δ30–36) RNA carried a 2′ deoxynucleotide at position G28 and was synthesized by Integrated DNA Technologies. All other RNAs were generated using *in vitro* transcription as previously described (14). Briefly, a synthetic oligonucleotide (Integrated DNA technologies) carrying the reverse complement of the desired sequence was annealed to a synthetic oligonucleotide corresponding to the T7 promoter to generate the transcription template. All transcribed RNAs carried two G nucleotides at their 5′ ends, carried over from the sequence of the T7 promoter. After 3–5 h, the reaction was treated with RQ1 DNase (Promega), and RNAs were purified by denaturing polyacrylamide gel electrophoresis. Purified RNAs were treated with calf-intestinal phosphatase and 5′-[^{32}P]-radiolabeled using T4 polynucleotide kinase and [γ-^{32}P] ATP, as previously described (14).

Electrophoretic mobility shift assays

TtCas6A and TtCas6B are positively charged at or near neutral pH (with predicted theoretical pI values of 9.3 and 9.7, respectively). To improve electrophoretic mobility on native polyacrylamide gels, MBP fusions of TtCas6A and TtCas6B (with predicted theoretical pI values of 6.1 and 6.6, respectively) were used in all binding assays. Protein concentrations were determined with a NanoDrop spectrophotometer (Thermo Scientific) using the calculated extinction coefficients at 280 nm for MBP–TtCas6A (96 260 M^{-1} cm^{-1}) and MBP–TtCas6B (93 280 M^{-1} cm^{-1}). To determine equilibrium product dissociation constants, 20 pM 5′–[^{32}P]-radiolabeled substrate RNA was titrated with increasing concentrations of protein in 20 mM HEPES (pH 7.5), 100 mM KCl, 5% glycerol, 0.01% Igepal-630, 1 mM DTT and 0.1 mg/ml yeast tRNA to prevent non-specific RNA binding. All binding reactions were incubated at 51°C for 10 min, as elevated temperatures were required for optimal RNA binding by TtCas6 proteins. Bound and unbound fractions were resolved on a native 8% polyacrylamide gel at room temperature using 1× TBE running buffer, detected by phosphorimaging using a Storm scanner (GE Healthcare) and quantified using ImageQuant software (GE Healthcare). Binding data were fit using Kaleidagraph (Synergy Software) according to the equation:

$$\text{fraction of substrate bound} = \text{A} \div (1+(K_d \div [Cas6])),$$

where A is the amplitude of the curve. The only exception was binding data collected for MBP–TtCas6A H37A using substrate R1, where a modified binding equation was applied using a Hill coefficient for negative cooperativity (n = 0.6). Binding data for the H37A mutant were plotted as follows:

$$\text{fraction of substrate bound} = \mathrm{A} \div ((1+(K_d \div [Cas6])))^{0,6},$$

with A being the curve amplitude. All reported $\mathrm{K_d}$ values represent arithmetic averages of three independent experiments, and error bars are reported as standard error of the mean (SEM).

Endonuclease cleavage assays

RNA cleavage assays were conducted under single-turnover conditions using 500-fold excess of enzyme over RNA substrate. All assays were performed at 51°C in 20 mM HEPES (pH 7.5), 100 mM KCl and 1 mM DTT. Cleavage reactions contained 10 nM substrate (0.5 nM of which was 5′-[^{32}P]-radiolabeled) and 5 μM protein. Ten-microliter aliquots were removed at indicated time points and quenched with 50 μl phenol:chloroform:isoamyl alcohol 25:24:1 (pH 8.0, Sigma Aldrich). Five microliters of the aqueous phase was subsequently resolved on a 15% denaturing (7 M urea) polyacrylamide gel. Cleaved and uncleaved RNA fractions were quantified by phosphorimaging as described earlier and fitted to a single exponential decay curve:

$$\text{fraction cleaved} = A \times (1 - \exp(-k \times t))$$

where A is the curve amplitude, k is the first-order rate constant and t is time. In the case of TtCas6A H37A, the curve amplitude (A) was fixed at 1 to avoid overestimating k, because the substrate was not cleaved to completion even at the latest time point. To reconstitute the activity of the catalytically impaired TtCas6A H37A mutant, the cleavage buffer was supplemented with 500 mM imidazole, pH 8.0.

To test whether TtCas6A and TtCas6B are single-turnover enzymes, cleavage assays were performed using 10 nM substrate (0.25 nM of which was 5′-[^{32}P]-radiolabeled) and varying molar concentrations of TtCas6A and TtCas6B (20 nM, 10 nM, 5 nM, 2.5 nM and 1.25 nM). Aliquots were taken at specified time points and the reactions were quenched using phenol/chloroform. The RNAs were subsequently resolved on a 15% denaturing (7 M urea) polyacrylamide gel and the cleaved and uncleaved fractions were quantified as described earlier.

Crystallization

Protein:RNA complexes were reconstituted at room temperature by incubating Cas6 with a 1.25-fold molar excess of RNA for 1 h in 20 mM HEPES (pH 7.5) and 250 mM KCl. The complexes were then purified by size-exclusion chromatography in the same buffer using a Superdex 75 (16/60) column (GE Life Sciences) and were subsequently concentrated to 5–10 mg/ml using 10 000 MWCO centrifugal concentrators (Millipore). Purified proteins and protein:RNA complexes were crystallized at 18°C using the hanging drop vapor diffusion method by mixing equal volumes of protein or protein–RNA complex and reservoir solution. TtCas6B crystallized in 0.1 M MES (pH 6.0), 0.2 M zinc acetate and 9% (w/v) PEG 6000. TtCas6A H37A crystallized in 0.1 M bis–tris propane (pH 6.5), 16% PEG 3350 and 0.2 M sodium sulfate. The TtCas6A–R1 substrate mimic complex was crystallized in 0.1 M bis–tris propane (pH 6.5), 18% (w/v) PEG 3350 and 0.3 M sodium sulfate. Crystals of TtCas6A–R1 product complex were obtained from 0.2 M sodium sulfate, 20% (w/v) PEG 3350. The TtCas6B–R3 product complex crystallized in 0.1 M Tris (pH 8.5), 13% (w/v) PEG 20 000. Typically, crystals formed within 1–3 days and were fully grown within 1 week.

Structure determination

All diffraction data were indexed, integrated and scaled using XDS (16). For TtCas6B, experimental phases were obtained from a three-wavelength multi-wavelength anomalous diffraction dataset collected at the Zn K-edge. The positions of zinc sites were determined using Phenix.hyss (17). Phases were subsequently calculated and improved by density modification using solvent flipping using AutoSharp (18). The resulting electron density maps were of excellent quality, and an initial model of TtCas6B could be readily built in COOT (19). Iterative cycles of model building in COOT and refinement in Phenix.refine (20) yielded a final model with R_{work} of 21.1% and R_{free} of 23.9%. The structure of TtCas6B bound to R3 RNA was determined by molecular replacement using the Phaser module in Phenix (21,22). The apo-TtCas6B structure was used as the search model, and the resulting phases gave electron density maps that revealed the presence of two RNA molecules bound to the TtCas6B dimer. These were initially built in COOT using idealized dsRNA. The structure was completed by iterative building and refinement cycles using COOT and Phenix.refine. Final refinement statistics for all models are shown in Supplementary Table S1.

All TtCas6A structures were solved by molecular replacement. Initially, the apo-TtCas6B atomic model was used to generate a search model for TtCas6A in Phenix. This model was subsequently used to solve the structure of the TtCas6A:R1 substrate RNA complex. Initial molecular replacement phases were improved using the prime-and-switch algorithm in RESOLVE (23) to produce a readily interpretable electron density map revealing two TtCas6A molecules and one RNA molecule in the asymmetric unit. The initial atomic model of TtCas6A was built automatically in Arp/wArp (24), while the RNA molecule was built in COOT starting from an idealized RNA hairpin structure. The TtCas6A H37A and TtCas6A–R1 product complex structures were solved by molecular replacement using the refined model of TtCas6A and the R1 RNA from the substrate complex. Atomic models were built using COOT and refined in Phenix.refine. All models have excellent stereochemistry, as judged using the Molprobity server (25), with more than 95% of all protein amino acid residues in the preferred regions of the Ramachandran plot and no Ramachandran outliers (Supplementary Table S1).

RESULTS

TtCas6A and TtCas6B bind and cleave CRISPR repeats R1 and R3 and retain their product RNAs after cleavage

The genome of *T. thermophilus* HB8 harbors 11 CRISPR loci containing three distinct types of repeats, termed R1-3 herein (Figure 1A; Supplementary Figure S1). All CRISPR loci are constitutively transcribed (26,27). Irrespective of the CRISPR locus of origin, all crRNAs in *T. thermophilus* contain a 5′-terminal eight-nucleotide handle derived from the repeat sequence that results from sequence-specific cleavage at the 3′ end of the hairpin structure predicted in each crRNA repeat (27). Three Cas6 genes have been identified in the *T. thermophilus* genome: TTHB231, TTHB192 and TTHA0078 (Supplementary Figure S1A). Previous structural and biochemical studies showed that the TTHB192 gene product, a member of the Cas6e subfamily, cleaves the R2 repeat found in the two spacer/repeat arrays flanking the Type I-E (*E. coli* subtype) Cas operon in the *T. thermophilus* genome (10,11). While TTHB231 is embedded in a hybrid Type I operon flanked by R3 repeat loci, TTHA0078 is not part of any CRISPR locus.

To determine whether the gene products of TTHA0078 and TTHB231 (hereafter referred to as TtCas6A and TtCas6B, respectively) are responsible for processing pre-crRNAs originating from R1 and/or R3 repeat loci, recombinant TtCas6A and TtCas6B proteins were expressed and purified from *E. coli* and tested for endonucleolytic activity using *in vitro* transcribed RNAs. Both proteins cleaved R1 and R3 repeat RNAs efficiently, whereas neither was able to cleave R2 repeat RNA (Supplementary Figure S1B). To characterize the binding affinities of TtCas6A and TtCas6B to their cognate crRNA repeats, we performed EMSAs using 5′-[^{32}P]-radiolabeled R1 and R3 repeat RNAs. The assays were carried out at 51°C to ensure RNA binding by the enzymes. As endonucleolytic cleavage occurred to completion during the course of the equilibrium binding reactions, the calculated equilibrium dissociation constants reflect product, rather than substrate, binding. Protein TtCas6A bound to the R1 repeat cleavage product with an apparent K_d of 90 ± 21 pM, whereas binding to the R3 repeat cleavage product was approximately 9-fold weaker (808 ± 154 pM). TtCas6B bound to R1 and R3 repeats with comparable dissociation constants of 1.96 ± 0.28 nM and 3.90 ± 0.78 nM, respectively (Figure 1B). The observed high-affinity product binding is consistent with the conclusion that, like many other Cas6 ribonucleases, both TtCas6A and TtCas6B function as single-turnover enzymes (8,10,14). To test this hypothesis, we performed cleavage assays at a range of substrate:enzyme molar ratios, measured the rate of cleavage and quantified the product yield (Figure 1C). The cleavage reaction yields scaled proportionally to enzyme concentration at sub-stoichiometric enzyme concentrations, while the apparent first-order rate constants remained essentially unchanged, indicating single-turnover catalysis. Collectively, these findings suggest that TtCas6A and TtCas6B are involved in processing precursor transcripts of repeat R1 and R3-containing CRISPR loci in *T. thermophilus* and that both enzymes likely remain bound to their products following cleavage.

Crystal structures of RNA-bound TtCas6A and TtCas6B

To determine how TtCas6A and TtCas6B bind and cleave their RNA substrates, we solved crystal structures of these proteins, both alone and in complexes with their cognate RNAs (Figure 2A, Supplementary Table S1). For TtCas6A, we obtained a structure of the enzyme bound to an RNA substrate mimic based on the R1 repeat sequence, consisting of the R1 stem-loop flanked by two additional nucleotides on either end of the stem. Cleavage of the substrate mimic RNA was prevented by introducing a 2′-deoxyribonucleotide at the G28 position, thereby removing the 2′-hydroxyl nucleophile required for the cleavage reaction. In addition to the substrate mimic complex, a crystal structure of a complex of TtCas6A and a cleaved RNA product was obtained when full-length R1 repeat was bound to wild-type TtCas6A and allowed to undergo cleavage during subsequent complex purification and crystallization. Finally, we determined the crystal structure of the TtCas6A H37A mutant, lacking a critical active-site residue, in the absence of bound RNA. For TtCas6B, we determined crystal structures of the wild-type enzyme alone and in complex with a product of the R3 repeat cleavage reaction.

Overall, both TtCas6A and TtCas6B adopt tandem ferredoxin/RRM folds similar to those observed for TtCas6e, PfCas6, *Sulfolobus solfataricus* Cas6 proteins SsCas6 and SsoCas6 (encoded by SSO2004 and SSO1437, respectively), as well as a non-catalytic Cas6 homolog from *Pyrococcus horikoshii* (Supplementary Figure S2A and B) (10,11,15,28–30). The N-terminal RRM domains of the two TtCas6 proteins also superimpose well with the single RRM fold found in the structure of PaCas6f (8). The two TtCas6 enzymes are highly similar to each other and superimpose with a root-mean-square deviation of 2.1 Å over 227 Cα atoms, reflecting the high degree of sequence identity (32%) between the two proteins (Supplementary Figure S2A and B).

In all crystal structures, both TtCas6A and TtCas6B form crystallographic (RNA-free TtCas6B) or non-crystallographic (all TtCas6A structures and the TtCas6B-product complex) dimers (Supplementary Figure S3A–C), consistent with size-exclusion chromatography results indicating that both enzymes are dimers in solution. The buried surface area of the TtCas6A dimer is 924 Å^2, whereas the TtCas6B dimer buries 1008 Å^2. Strikingly, while the TtCas6B-R3 product crystal structure reveals a 2:2 stoichiometry, both substrate and product TtCas6A–R1 RNA complexes crystallized with an apparent 2:1 stoichiometry, with only one RNA molecule bound to the non-crystallographic TtCas6A dimer. Size-exclusion chromatography of TtCas6A and TtCas6B-RNA complexes used for crystallization as well as their absorbance ratios at 280 and 260 nm were indicative of 2:2 stoichiometry (data not shown). Additionally, both proteins behaved similarly in cleavage assays (Figure 1C), and no negative cooperativity was observed for TtCas6A in binding assays. The apparent 2:1 stoichiometry of the TtCas6A–RNA complexes is

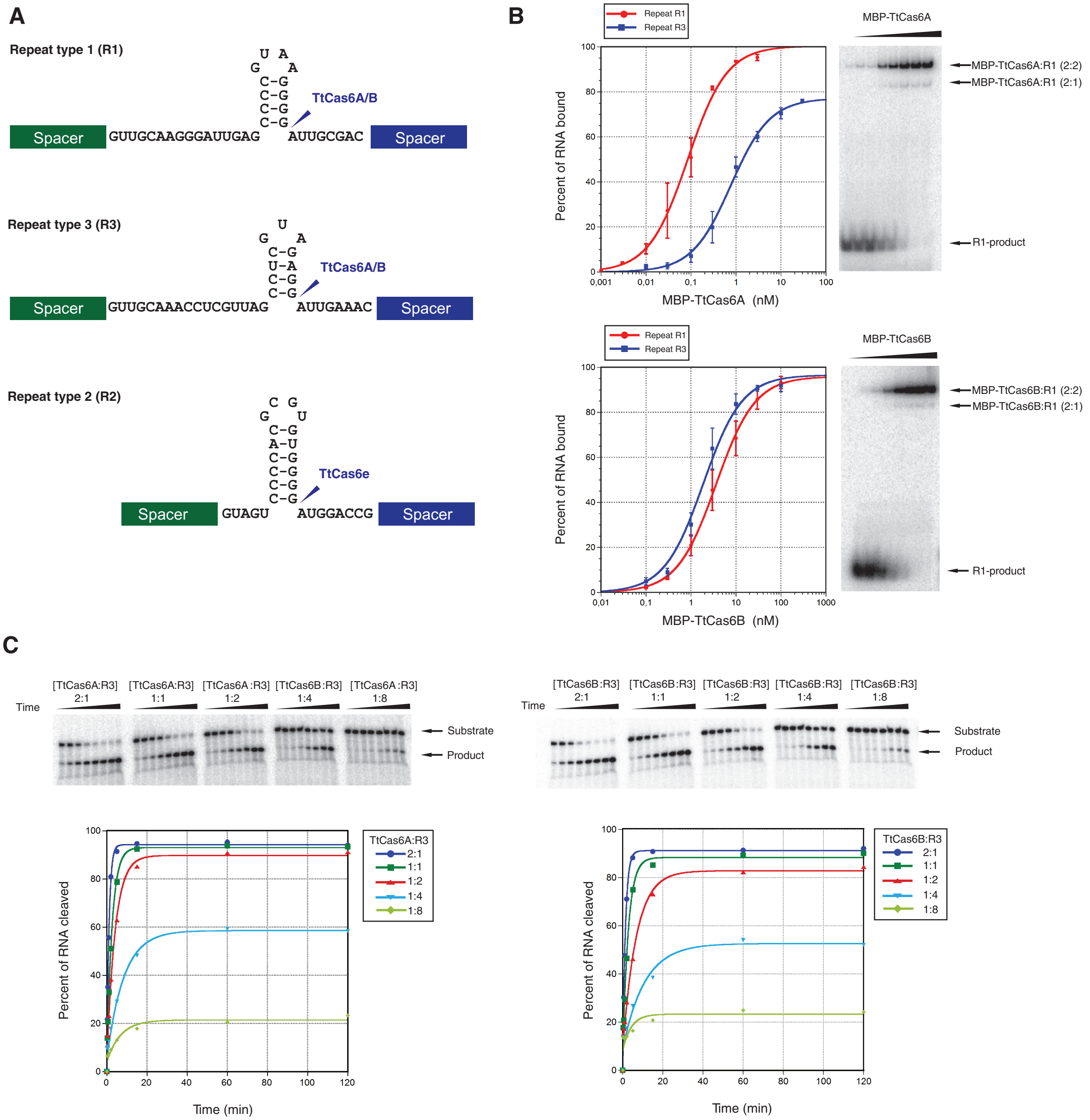
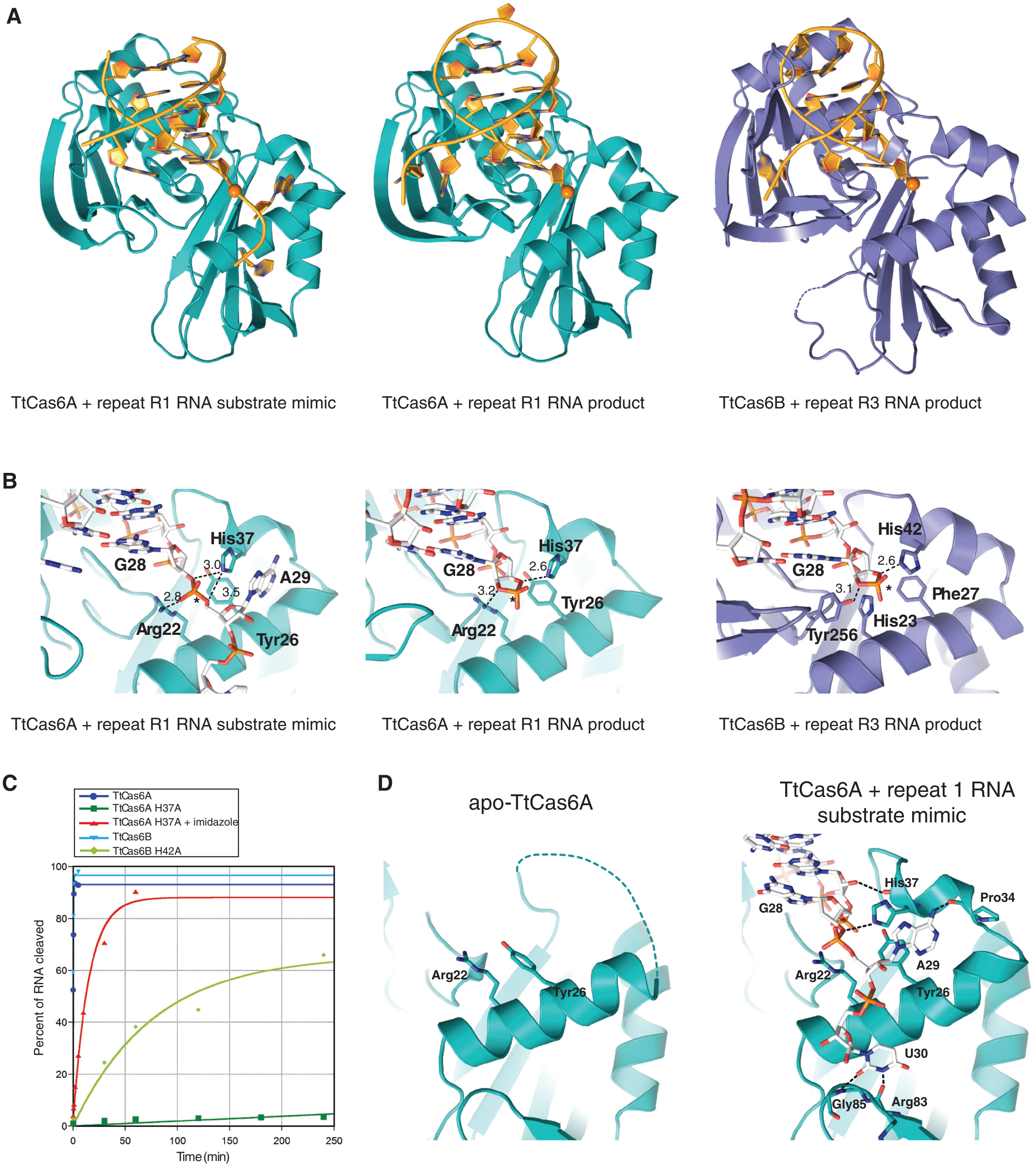


Figure 1. TtCas6A and TtCas6B both cleave repeats R1 and R3 and retain their cleaved products. (**A**) Sequences and predicted secondary structures of *T. thermophilus* CRISPR repeats. Sites of cleavage are indicated with blue arrows. TtCas6e (TTHB192) cleaves repeat R2, while TtCas6A (TTHA0078) and TtCas6B (TTHB231) both cleave repeats R1 and R3. (**B**) Cleavage product binding affinities of TtCas6A and TtCas6B enzymes. Maltose-binding protein (MBP)-fused TtCas6A or TtCas6B were bound to 5′-[^{32}P]-radiolabeled, *in vitro* transcribed R1 and R3 RNAs. Bound and unbound fractions were resolved by electrophoresis on a native polyacrylamide gel and visualized by phosphorimaging. The data for these and all subsequent binding assays were fit with standard binding isotherms (solid line), unless otherwise stated. Error bars on each data point denote standard error of the mean (SEM) from three independent experiments. (**C**) Kinetics experiments to confirm single turnover. RNA cleavage assays were carried out at indicated protein:RNA ratios. RNA cleavage was monitored using denaturing polyacrylamide gel electrophoresis. The data from these and all subsequent endoribonuclease activity assays were fit with single exponential curves to yield first-order rate constants.

therefore likely a crystallization-induced artifact. The dimer interfaces are highly similar to those observed in the structures of *S. solfataricus* Cas6 proteins (SsCas6 and SsoCas6) (29,30), suggesting that dimer formation is a general property of many Cas6 proteins (Supplementary Figure S3D).

Mechanism of substrate recognition and cleavage

As anticipated, the R1 and R3 repeat RNAs form stem-loop structures. In both proteins, the RNAs bind in a positively charged cleft located between the two RRM folds, as observed in the structures of TtCas6e–R2



TtCas6A + repeat R1 RNA substrate mimic

TtCas6A + repeat R1 RNA product

TtCas6B + repeat R3 RNA product

TtCas6A + repeat R1 RNA substrate mimic

TtCas6A + repeat R1 RNA product

TtCas6B + repeat R3 RNA product

apo-TtCas6A

TtCas6A + repeat 1 RNA substrate mimic

Figure 2. Structures of TtCas6A and TtCas6B enzymes bound to substrate mimic and product RNAs. (**A**) Ribbon diagrams showing the overall views of Cas6–RNA complexes: TtCas6A–R1 substrate mimic (left), TtCas6A–R1 product (middle) and TtCas6B–R3 product (right). Bound RNAs are depicted in cartoon format and colored in yellow. The scissile phosphate groups are depicted as orange spheres. All cartoon molecular diagrams were generated using Pymol (http://www.pymol.org). (**B**) Zoomed-in views of the TtCas6 active sites, shown in the same orientation as in A. Hydrogen-bonding interactions are denoted with dashed lines; numbers indicate interatomic distances in Å. (**C**) Endonuclease activity assays of wild-type (WT) and active-site mutant proteins. For the TtCas6A H37A mutant, the cleavage assay was additionally carried out in the presence of 500 mM imidazole. (**D**) Active site of TtCas6A undergoes conformational ordering on substrate recognition. Left: zoomed-in view of the active site in the RNA-free TtCas6A molecule in the 2:1 protein–R1 substrate mimic complex. Right: zoomed-in view of the active site in the RNA-bound TtCas6A molecule. Hydrogen-bonding interactions are denoted with dashed lines.

repeat complexes (10,11). In further analogy with TtCas6e, TtCas6A and TtCas6B complexes also insert a beta-hairpin from their C-terminal RRM domains into the major groove of the dsRNA stems (Figure 2A). In the TtCas6A–substrate mimic complex, the two nucleotides downstream of the scissile phosphate are recognized in a sequence-specific manner through base-specific interactions (described in detail later). The structures of the RNA product complexes of both TtCas6A and TtCas6B reveal 2′–3′ cyclic phosphate groups in the respective active sites, consistent with a catalytic mechanism involving nucleophilic attack by the 2′-hydroxyl of the upstream nucleotide (G28) (Figure 2A and B).

The active site of TtCas6A is located in a pocket surrounded by helix α1 and the α1-β2, β10-β11 and α5-β12 loops (Figure 2B). The scissile phosphate group is contacted by Arg22 and His37, and positioned in an extended conformation that would permit an in-line attack by the 2′-hydroxyl of G28 (Figure 2B). His37 is positioned to hydrogen bond with the 5′ or 3′ bridging oxygen atoms, and might therefore act as the general acid that protonates the leaving group during catalysis, in addition to charge-stabilizing the scissile phosphate. The active site of TtCas6B is composed of His23, His42 and Tyr256, whereby Tyr256 and His42 hydrogen bond to the 2′ and 3′ oxygens of the cyclic phosphate product. In a substrate complex, Tyr256 would likely be positioned to deprotonate the 2′-hydroxyl of G28 during nucleophilic attack, while His42, in analogy with His37 in TtCas6A, would stabilize the scissile phosphate and protonate the leaving group.

To shed light on the catalytic mechanism of TtCas6A and TtCas6B, we performed cleavage assays using wild-type proteins as well as active-site mutants TtCas6A H37A and TtCas6B H42A using repeat R1 as a substrate (Figure 2C). The first-order rate constants determined under single-turnover conditions for wild-type TtCas6A ($3.2\,\mathrm{min}^{-1}$) and TtCas6B ($3.7\,\mathrm{min}^{-1}$) are in good agreement with first-order rate constants previously determined for TtCas6e and PaCas6f (10,14). Strikingly, we found TtCas6A H37A to be almost inactive (~17 000-fold cleavage defect), whereas TtCas6B H42A showed only a ~300-fold cleavage defect, indicating that despite considerable structural homology, the catalytic mechanisms of TtCas6A and TtCas6B might be substantially different. To confirm the role of the active-site histidine His37 in TtCas6A, we sought to replace the histidine side chain by adding imidazole (a histidine mimic) to the cleavage reaction. This protein complementation strategy using imidazole has been used recently to convert PaCas6f into an inducible endoribonuclease (31). In the presence of 500 mM imidazole, the cleavage rate of TtCas6A H37A was enhanced ~360-fold, underscoring the importance of the active-site histidine in the catalytic mechanism of TtCas6A.

In contrast to TtCas6A and TtCas6B and many other Cas6 enzymes, SsCas6 and SsoCas6 (both of which are active ribonucleases) lack histidine residues at the position equivalent to H37 in TtCas6A, suggesting that conserved lysines (Lys25 and Lys28) in helix α1 in both SsCas6 and SsoCas6 could act as the key catalytic residues instead (29,30). To determine whether the equivalent and highly conserved residues in TtCas6A (Arg22) and TtCas6B (His23) are also involved in catalysis, we mutated these residues to alanine and performed cleavage experiments. The resulting first-order rate constants were less than 7-fold lower relative to the wild-type proteins (Supplementary Figure S4), suggesting that these residues contribute to substrate binding and stabilization of the scissile phosphate group during cleavage, but they are unlikely to function as general acid or base catalysts in the chemistry of RNA cleavage.

The active site of TtCas6A undergoes a conformational ordering on RNA binding

The crystal structures of TtCas6A–RNA complexes allow comparisons of the RNA-free and RNA-bound states of the enzyme due to the presence of an RNA-free TtCas6 molecule in the crystallographic asymmetric unit. In the RNA-free TtCas6A, the loop connecting helix α1 and strand β2 (residues 33–40), which contains the active-site histidine His37, is disordered (Figure 2D). On substrate RNA binding, the loop becomes ordered and forms a short helical segment, as the backbone carbonyls of Pro40 and His37 form hydrogen bonds with the 2′ hydroxyl groups of G26 and G27, respectively, and the His37 side chain forms a hydrogen bond with the 3′-hydroxyl oxygen of G28. Additional interactions mediate substrate recognition downstream of the scissile phosphate; the 6-amino group of A29 forms a hydrogen bond with the amide carbonyl of Pro34, while U30 is specifically recognized through hydrogen-bonding interactions with the backbone amide of Gly85 and carbonyl of Arg83. The ordering of the His37-containing active-site loop persists in the product complex, suggesting that scissile phosphate recognition by His37 and additional interactions with the ribose–phosphate backbone upstream of the cleavage site drive the conformational change on substrate binding.

Recognition of RNA sequence and geometry by TtCas6A and TtCas6B

In both TtCas6A and TtCas6B, extensive networks of ionic and hydrogen-bonding interactions are involved in RNA recognition (Figure 3A and B). In both proteins, the RNA stem-loop straddles the β10–β11 loop, and is positioned in a cleft between the active-site loop and a beta-hairpin (β7–β8) that inserts into the major groove. In TtCas6A, the hairpin presents Arg129 for sequence-specific hydrogen-bonding contacts with the lower three C-G base pairs in the stem (Figure 3A). TtCas6B lacks an equivalent residue in the major groove-binding hairpin. Instead, the side chain of Ser147 hydrogen-bonds to the base of G25, as the only sequence-specific contacts with the RNA (Figure 3A). In both Cas6–RNA complexes, the ribose–phosphate backbone in the 3′ half of the stem-loop is anchored through a series of hydrogen-bonding contacts involving the phosphate groups of nucleotides 25–28 and the 2′-hydroxyl groups of nucleotide G26 in

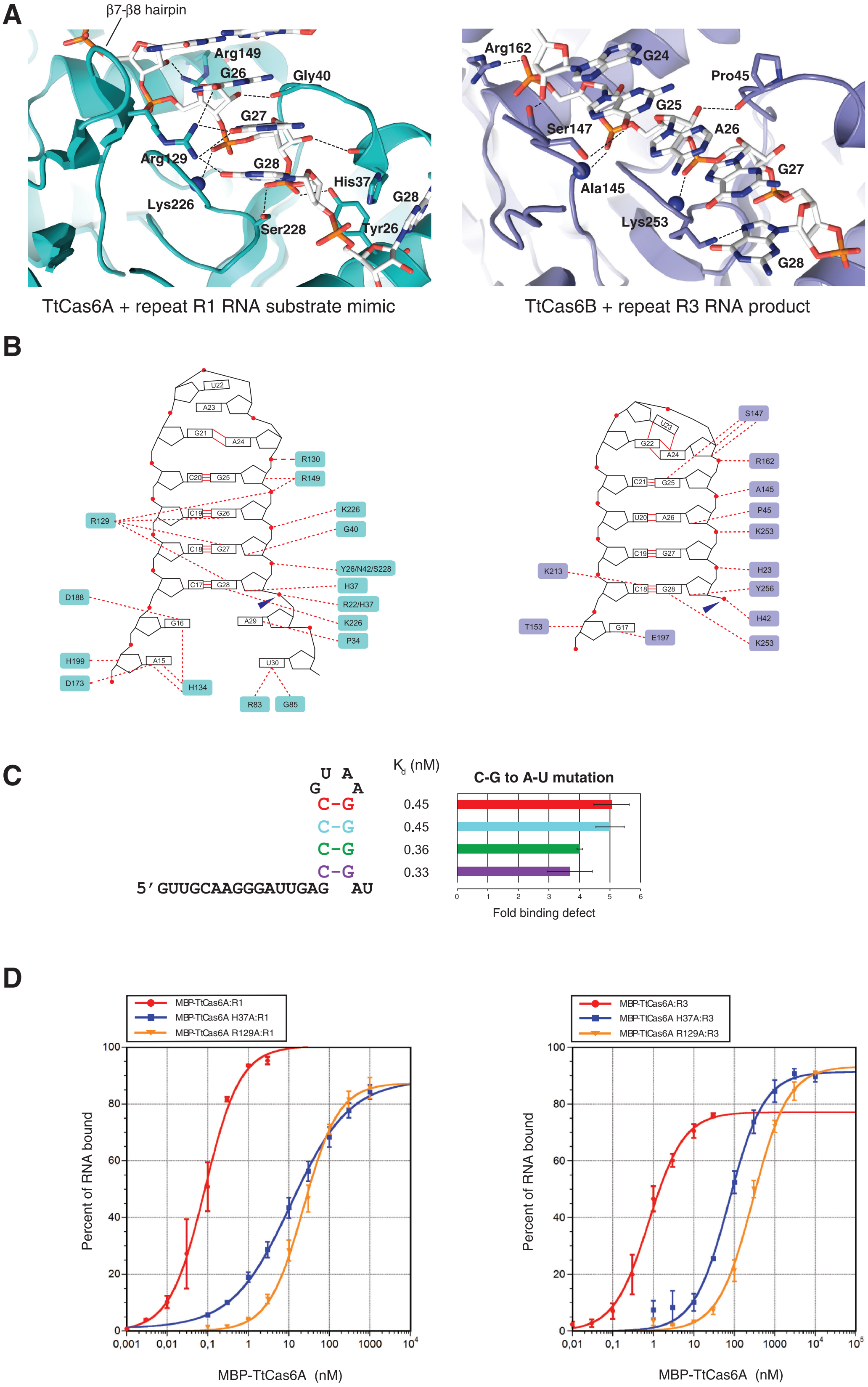


Figure 3. RNA recognition by TtCas6A and TtCas6B. (**A**) Detailed views of RNA binding by TtCas6A (left) and TtCas6B (right). Hydrogen-bonding interactions are indicated with black dashed lines. Blue spheres denote backbone amide nitrogen atoms of Lys226 in TtCas6A (left) and Ala145 and Lys253 in TtCas6B (right). (**B**) Schematic diagrams of protein–RNA contacts in the TtCas6A–R1 substrate mimic (left) and TtCas6B–R3 product complexes. Amino acid residues contacting the bound RNA via ionic or hydrogen-bonding interactions are highlighted. Blue arrows mark the scissile phosphates. Red circles denote phosphodiester groups in the RNA backbone. Red lines indicate base-pairing interactions. (**C**) Base-pair

(continued)

the TtCas6A–RNA complexes and nucleotide G27 in the TtCas6B–product complex, respectively (Figure 3A and B). A conserved lysine residue (Lys226 in TtCas6A and Lys253 in TtCas6B) found in the α5-β12 loop contacts the base of G28 with its side chain, while making a hydrogen-bonding interaction between its backbone amide and the phosphate group linking nucleotides G26 and G27. The α5-β12 loop, which corresponds to the Gly-rich loop, a notable feature of the RAMP superfamily (32,33), makes additional interactions with the RNA substrates through Ser228 in TtCas6A and Tyr256 in TtCas6B. Together, at least five residues within the beta-hairpin and GhGxxxxGhG motifs interact with the RNA substrates (Arg129, Agr130, His134, Lys226 and Ser228 in TtCas6A; Ala145, Ser147, Thr153, Lys253 and Tyr256 in TtCas6B) and about half of them do so in a base-specific manner (Figure 3A and B). Therefore, as in the structures of PaCas6f (PaCsy4) and TtCas6e (TtCse3), the RNAs are recognized both via their sequence and their shape (8,10,11).

To test the importance of the stem sequence for substrate RNA recognition by TtCas6A, a series of EMSAs were performed using R1 repeat-derived RNAs that carried single base-pair substitutions (C-G → A-U). All mutant RNAs contained the complete 5′ segment and additional two nucleotides downstream of the cleavage site. Compared with the wild-type RNA, substitution of any of the four C-G base pairs in the stem led to about 5-fold decrease in affinity (Figure 3C). This is consistent with the observation that the lower three C-G base pairs are specifically read out by Arg129. The binding defect observed on mutation of the closing (uppermost) base pair could be due to destabilization of the stem-loop structure, as loop stability is typically governed by the closing base pair (34). To further investigate the protein determinants of RNA binding, we tested the TtCas6A mutants H37A and R129A and performed binding assays using substrates R1 and R3 (Figure 3D). Mutation of Arg129 resulted in a strong binding defect with ~260- and ~290-fold decrease in affinity for R1 and R3, respectively, when compared with wild-type TtCas6A, in agreement with the observed function of this residue in simultaneous recognition of the lower three C-G base pairs in the RNA stem-loop. A similar, but somewhat weaker, effect was observed for TtCas6A H37A, which yielded ~70- and ~90-fold reduction in affinities for R1 and R3 repeat RNAs, respectively. The binding defects indicate that besides playing a key role in the catalysis of RNA cleavage, the His37 side chain also contributes to substrate binding. This is consistent with the ordering of the active-site loop observed on substrate recognition, which appears to be driven in part by the interaction between the His37 side chain and the scissile phosphate group.

Recognition of the unstructured 5′-segment of the repeat suggests a two-site model for RNA binding

The structures of TtCas6A– and TtCas6B–product RNA complexes reveal that besides recognizing the stem-loop, the enzymes also make specific interactions with the upstream RNA sequence. In the TtCas6A–R1 product complex, two nucleotides upstream of the stem-loop are observed in $2F_o$-F_c electron density maps. The remainder of the 5′ segment of the R1 repeat RNA is not ordered, although the RNA is intact in the crystal (data not shown). The purine bases of the two ordered nucleotides in the 5′ segment are inserted into a crevice at the interface of the two TtCas6A molecules in the non-crystallographic dimer (Figure 4A). G16 engages in hydrogen bonding with the side chain of His134 and the backbone carbonyl of Asp188. The base of A15 is hydrogen bonded to the backbone amide and carbonyl groups of His134. In the TtCas6B–R3 product complex, the two RNA molecules in the asymmetric unit adopt slightly different conformations at their 5′ ends. In one molecule, only one nucleotide (G17) upstream of the repeat stem-loop is ordered, forming hydrogen bonds with the side chains of Arg208 and Glu197, each contributed by one TtCas6B molecule in the non-crystallographic dimer (Figure 4A). In the other RNA molecule, both G17 and A16 are ordered, and the base of A16 is tucked in and stacks below the terminal base pair of the R3 repeat stem-loop (Supplementary Figure S5A).

Inspection of the molecular surface of TtCas6A reveals a deep groove tracing the junction of the two RRM folds. This groove extends from the A15-binding site towards a highly positively charged patch located on the reverse side of the protein from the active site (Figure 4B). A similar groove is observed in TtCas6B (Supplementary Figure S5B). A sulfate ion is bound to the basic patch in the structures of both TtCas6A–R1 substrate and TtCas6A–R1 product complexes, and is contacted by the side chains of Arg121 and Arg223 (Supplementary Figure S5C). In the structure of PfCas6 bound to a fragment of its cognate repeat RNA, nucleotides 2–10 of the repeat bind to a positively charged groove located on the face of the protein opposite from the active site (15). Superposition of the PfCas6 and TtCas6A RNA complex structures reveals that the basic groove in TtCas6A overlaps with the PfCas6A RNA binding site such that the 3′ end of the bound PfCas6 RNA fragment (nucleotide A10) aligns with the 5′ end (nucleotide A15) of the R1 repeat RNA (Figure 4C). This suggests that the basic groove in TtCas6A might constitute an additional RNA binding site that interacts with the unstructured 5′ segment of the R1 repeat RNA upstream of A15. Although neither nucleotides G1-G14 of the R1 product RNA nor nucleotides G1-U15 of the R3

Figure 3. Continued
contributions to R1 repeat recognition by TtCas6A. A series of RNAs in which individual C-G base pairs were substituted with A-U were prepared and assayed for binding to TtCas6A using EMSAs. The data for each base-pair substitution are expressed as K_d and as fold reduction in affinity relative to wild-type R1 RNA. The color-coding follows the schematic diagram of the R1 RNA (left). (**D**) R1 (left) or R3 (right) product RNA binding by WT TtCas6A, R129A or H37A mutants was quantified using EMSAs. The data are plotted as in Figure 1B, with the exception of TtCas6A H37A, for which a modified equation using a Hill coefficient for negative cooperativity (n = 0.6) was used.

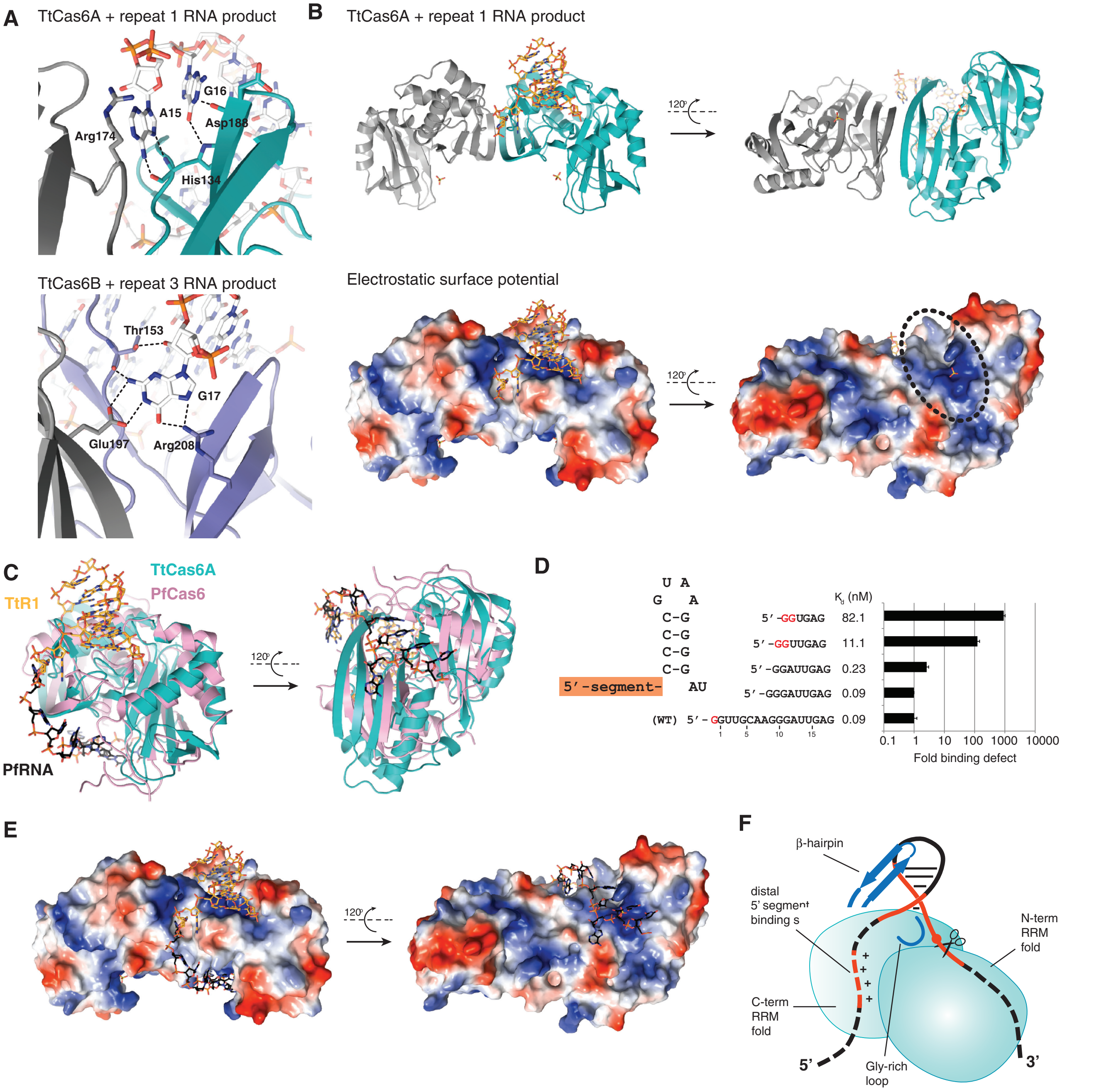


Figure 4. Recognition of the 5′ segment of the repeat RNA. (**A**) Details of sequence-specific recognition of nucleotides upstream of the stem-loop in RNA repeats. Top: TtCas6A–R1 product complex. Nucleotides 1–14 of the R1 product RNA are disordered. Bottom: TtCas6B–R3 product complex. Nucleotides 1–15 of the R3 product RNA are disordered. (**B**) Surface electrostatic potential map of TtCas6A identifies a second RNA binding site. Top: Cartoon diagram of the 2:1 TtCas6A–R1 product RNA complex. RNA is shown in orange. Bound sulfate ions are depicted in stick format. Bottom: Electrostatic surface potential map of TtCas6A, shown in the same orientations as earlier. Blue, positively charged region; red, negatively charged region. The positively charged patch located on the surface opposite from the active site is highlighted with a black ellipse. (**C**) Structural superposition of the TtCas6A–R1 product RNA (TtR1) and PfCas6–repeat RNA (PfRNA) (PDB code: 3PKM) complexes. TtCas6A is colored teal; PfCas6 is colored pink. *T. thermophilus* R1 repeat RNA is colored orange. PfRNA is colored black. Nucleotide A15 of TtR1 aligns with G10 of PfRNA. (**D**) Nucleotides in the single-stranded 5′ segment of R1 repeat RNA contribute to binding. TtCas6A binding to a series of truncated RNAs based on the R1 repeat was quantified by EMSAs as in Figure 1B. The data are expressed as K_d and as a fold binding defect relative to wild-type R1 repeat. 5′-terminal G nucleotides resulting from *in vitro* transcription are shown in G. (**E**) Structural superposition of TtCas6A dimer with *P. furiosus* repeat RNA, based on the superposition shown in D. TtCas6A is colored according to surface electrostatic potential and shown in the same orientations as in B. TtR1 RNA is colored orange; PfRNA is colored black. (**F**) Cartoon model of RNA recognition by TtCas6 enzymes. TtCas6A binds the stem-loop region of the RNA (red solid line) at the interface of the two RRM-like domains. The two major elements responsible for the interaction are the variable beta-hairpin and the Gly-rich loop (both depicted in blue). Additionally, the 5′ segment of the repeat RNA (dashed red line) is bound by a distal positively charged cleft.

RNA are ordered in the respective co-crystal structures of TtCas6A and TtCas6B, it is possible that in both cases, this is a consequence of the high ionic strength of the crystallization condition and the presence of sulfate in the TtCas6A crystals.

We therefore hypothesized that the unstructured 5′ segment contributes to R1 RNA binding by TtCas6A. To test this, we measured the affinity of TtCas6A for a series of RNAs based on the R1 repeat in which nucleotides were progressively removed from the 5′ end (Figure 4D). Deletion of nucleotides 1–8 had little effect on binding affinity. In contrast, truncation of the R1 repeat RNA beyond nucleotide G9 led to a gradual loss of binding affinity, with an approximately 900-fold increase in K_d on deletion of residues 1–13 of the R1 repeat RNA and a complete loss of binding on deletion of nucleotides 1–14 (Figure 4D). This indicates that nucleotides 9–13 of the R1 repeat RNA contribute ~4 kcal.mol^{-1} to the binding free energy. Together, these findings suggest that the 5′ segment of the repeat RNA is recognized by TtCas6A, in a sequence-specific manner, and hint at the existence of a second RNA binding site in TtCas6A that specifically interacts with nucleotides 9–13 of the repeat.

DISCUSSION

Cas6 enzymes constitute a class of highly sequence- and structure-specific endoribonucleases responsible for the maturation of crRNAs in Type I and III CRISPR systems. These proteins constitute a clade within the larger RAMP superfamily (6,32,33,35). Previous structural studies showed that several members of this enzyme class share a common ferredoxin/RRM fold that provides a platform for pre-crRNA binding and endonucleolytic processing. However, the extreme sequence diversity, differences in active-site architecture and distinct RNA binding modes recognizing RNA hairpin substrates in some cases and single-stranded substrates in others have made it difficult to understand how Cas6 enzymes might have evolved from a common ancestral protein.

To determine the evolutionary relationship between distinct members of the Cas6 clade, we determined crystal structures of two Cas6 enzymes involved in crRNA processing both alone and in complexes with substrate and product RNAs. These structures suggest how the RRM protein scaffold common to Cas6 endonucleases has evolved to recognize two distinct RNA structural features. In several respects, the RNA binding mode observed in TtCas6A and TtCas6B resembles that identified previously for Cas6e and Cas6f enzymes. In all of these ribonucleases, the terminal base pair at the bottom of an RNA hairpin substrate straddles a beta-hairpin in the C-terminal RRM domain. A highly variable region inserted between the first beta-strand and the first α-helix of the second RRM domain forms a secondary structure motif (beta hairpin in TtCas6e, TtCas6A and TtCas6B or alpha helix in PaCas6f) that probes the major groove of the RNA to provide sequence- and shape-specific readout (8,10,11). Another major determinant of RNA recognition is the Gly-rich loop motif (GhGxxxxGhG, where h stands for a hydrophobic residue and xxxxx contains at least one arginine or lysine), a signature feature of the RAMP superfamily (32,33). In TtCas6A and TtCas6B, as well as in other Cas6 enzymes (TtCas6e and SsoCas6), this loop contributes to RNA recognition by providing multiple contacts to the phosphate backbone of the substrate RNAs. The highly divergent sequences found in the beta-hairpin and GhGxxxxGhG motifs are therefore the major sources of variability in RNA binding observed across diverse Cas6 enzymes. Together, the two motifs have thus provided a scaffold for the evolution of diverse RNA binding modes in Cas6 ribonucleases. Notably, in TtCas6B, the GhGxxxGhG motif also contacts the 2′-hydroxyl of the ribose immediately upstream of the scissile phosphate, suggesting that this motif may also play a previously unrecognized role in catalyzing RNA cleavage in a subset of Cas6 enzymes.

Our structural and biochemical results further reveal that RNA binding in both TtCas6A and TtCas6B involves additional interactions between the enzyme and nucleotides upstream of the stem-loop structure in the RNA. In the structures of both TtCas6A–R1 and TtCas6B–R3 product complexes, two unpaired nucleotides upstream of the stem-loop insert into a groove at the Cas6 dimer interface for sequence-specific recognition. This is reminiscent of RNA binding observed in a recent structure of SsCas6 (SSO2004) in complex with a CRISPR repeat RNA predicted to lack secondary structures (29). Here, SsCas6 specifically interacts with a three-base-pair stem-loop motif in the RNA as well as with three nucleotides upstream of the stem, which bind at the interface of the SsCas6 dimer. However, in contrast to SsCas6, our biochemical analysis of RNA binding by TtCas6A suggests that nucleotides further upstream in the unstructured 5′ segment (at positions 9–13 of the R1 repeat RNA) also make a substantial contribution to the overall affinity. This suggests that these nucleotides are also recognized in a sequence-dependent manner, even though the interaction may be transient and was therefore not captured in our crystal structures. We hypothesize that this interaction resembles in part the RNA recognition mode of PfCas6, a Cas6 protein that recognizes CRISPR repeat RNAs lacking hairpin secondary structures. PfCas6 binds nucleotides in the 5′-terminal sequence of its RNA substrate at a region distant from the endonuclease active site (15). Superposition of TtCas6A and PfCas6 crystal structures shows that the 5′ end of the hairpin RNA bound to TtCas6A aligns almost perfectly with the 3′ end of the single-stranded RNA bound to PfCas6 (Figure 4C,E). This remarkable alignment immediately suggests a model in which Cas6 proteins can provide two binding surfaces with complementary but orthogonal RNA recognition modes (Figure 4F). The two RNA binding sites have emerged and to some extent co-evolved in Cas6 enzymes, providing further plasticity to the substrate recognition mechanism and enabling these enzymes to accommodate an even wider variety of CRISPR repeat sequences and structures in different CRISPR–Cas systems.

A common property of Cas6 endonucleases is their high affinity for cleaved products and the resulting lack of multiple turnover (10,14). In a number of Type I CRISPR systems, the Cas6 endonuclease is an integral subunit of the targeting complex, as exemplified by the Cascade and Csy complexes (5,36,37). Here, retention of the cleaved RNA product by the Cas6 enzyme is thought to mediate the assembly of the targeting machineries. However, in Type III (the Cmr complex) as well as in some Type I systems (the archaeal Csa complex in *S. solfataricus*), the processing endonuclease does not stably associate with the targeting complex (38,39). Whether product binding by Cas6 is required for downstream steps in the interference mechanism of these CRISPR systems or whether it is simply a consequence of the highly selective RNA binding mechanism awaits further study.

The active sites of Cas6 enzymes also display remarkable plasticity. PaCas6f uses a catalytic dyad consisting of a histidine and a serine; TtCas6e uses a histidine, tyrosine and lysine, whereas PfCas6 contains a histidine and a tyrosine (7,8,10,11,40). Our structures of TtCas6A and TtCas6B underscore the near-universal occurrence of catalytic histidines in Cas6 enzymes. However, the histidine residues are not conserved in their position relative to the scissile phosphate and consequently play seemingly different roles in the catalytic mechanisms—deprotonating the attacking nucleophile in PaCas6f, protonating the leaving group in TtCas6e, TtCas6A, TtCas6B and PfCas6 or charge-stabilizing the scissile phosphate (His23 in TtCas6B) (10,15,40). In contrast, a subset of Cas6 enzymes (notably SsCas6 and SsoCas6) lack a histidine in their active sites, suggesting that other residues (notably lysines) in the alpha-helical segment C-downstream of the first beta-strand of the first RRM fold might assume catalytic roles instead (29,30). A parsimonious scenario for the evolution of Cas6 ribonucleases suggests that these proteins derived from an ancestral RNA-binding RAMP that was probably an active ribonuclease containing an active-site histidine residue (32). However, given that the catalytic efficiency of Cas6 enzymes is generally poor and the active-site architectures are highly variable across the family, the mechanisms of RNA cleavage have evolved and diversified dramatically since the last common ancestor of Cas6 ribonucleases. This may have been dictated at least in part by the precise structural requirements for specific RNA recognition.

ACCESSION NUMBERS

Atomic coordinates and structure factor amplitudes have been deposited in the Protein Data Bank for each of the structures listed: TtCasA-R1 substrate mimic complex (4C8Y), TtCas6A-R1 product complex (4C8Z), apo-TtCas6A H37A (4C97), apo-TtCas6B (4C98) and TtCas6B-R3 product complex (4C9D).

SUPPLEMENTARY DATA

Supplementary data are available at NAR Online, including [41–44].

ACKNOWLEDGEMENTS

The authors thank Samuel Sternberg for assistance with functional assays and data analysis. They thank members of the Doudna group for discussions and critical reading of the manuscript. They are grateful to Alison Smith for technical help. They thank James Holton and Corie Ralston (Lawrence Berkeley National Laboratory) for assistance with X-ray data collection. JAD is an Investigator of the Howard Hughes Medical Institute.

FUNDING

This work was supported by the Howard Hughes Medical Institute. Funding for open access charge: Howard Hughes Medical Institute.

Conflict of interest statement. None declared.

REFERENCES

1. Wiedenheft,B., Sternberg,S.H. and Doudna,J.A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, **482**, 331–338.
2. Marraffini,L.A. and Sontheimer,E.J. (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.*, **11**, 181–190.
3. Al-Attar,S., Westra,E.R., van der Oost,J. and Brouns,S.J.J. (2011) Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes. *Biol. Chem.*, **392**, 277–289.
4. Terns,M.P. and Terns,R.M. (2011) CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.*, **14**, 321–327.
5. Brouns,S.J.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J.H., Snijders,A.P.L., Dickman,M.J., Makarova,K.S., Koonin,E.V. and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
6. Makarova,K.S., Half,D.H., Barrangou,R., Brouns,S.J., Charpentier,E., Horvath,P., Moineau,S., Mojica,F.J.M., Wolf,Y.I., Yakunin,A.F. *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.*, **9**, 467–477.
7. Carte,J., Wang,R., Li,H., Terns,R.M. and Terns,M.P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.*, **22**, 3489–3496.
8. Haurwitz,R.E., Jínek,M., Wiedenheft,B., Zhou,K. and Doudna,J.A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*, **329**, 1355–1358.
9. Jore,M.M., Lundgren,M., van Duijn,E., Bultema,J.B., Westra,E.R., Waghmare,S.P., Wiedenheft,B., Pul,U., Wurm,R., Wagner,R. *et al.* (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.*, **18**, 529–536.
10. Sashital,D.G., Jínek,M. and Doudna,J.A. (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat. Struct. Mol. Biol.*, **18**, 680–687.
11. Gesner,E.M., Schellenberg,M.J., Garside,E.L., George,M.M. and Macmillan,A.M. (2011) Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat. Struct. Mol. Biol.*, **18**, 688–692.
12. Ebihara,A., Yao,M., Masui,R., Tanaka,I., Yokoyama,S. and Kuramitsu,S. (2006) Crystal structure of hypothetical protein TTHB192 from Thermus thermophilus HB8 reveals a new protein

family with an RNA recognition motif-like domain. *Protein Sci.*, **15**, 1494–1499.

13. Kunin,V., Sorek,R. and Hugenholtz,P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.*, **8**, R61.
14. Sternberg,S.H., Haurwitz,R.E. and Doudna,J.A. (2012) Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA*, **18**, 661–672.
15. Wang,R., Preamplume,G., Terns,M.P., Terns,R.M. and Li,H. (2011) Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure*, **19**, 257–264.
16. Kabsch,W. (2010) XDS. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
17. Zwart,P.H., Afonine,P.V., Grosse-Kunstleve,R.W., Hung,L.-W., Ioerger,T.R., McCoy,A.J., McKee,E., Moriarty,N.W., Read,R.J., Sacchettini,J.C. *et al.* (2008) Automated structure solution with the PHENIX suite. *Methods Mol. Biol.*, **426**, 419–435.
18. Vonrhein,C., Blanc,E., Roversi,P. and Bricogne,G. (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.*, **364**, 215–230.
19. Emsley,P., Emsley,P., Cowtan,K. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
20. Afonine,P.V., Grosse-Kunstleve,R.W., Echols,N., Headd,J.J., Moriarty,N.W., Mustyakimov,M., Terwilliger,T.C., Urzhumtsev,A., Zwart,P.H. and Adams,P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 352–367.
21. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phasercrystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
22. Adams,P.D., Afonine,P.V., Bunkóczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.-W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
23. Terwilliger,T.C. (2004) Using prime-and-switch phasing to reduce model bias in molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2144–2149.
24. Morris,R.J., Perrakis,A. and Lamzin,V.S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Meth Enzymol.*, **374**, 229–244.
25. Davis,I.W., Leaver-Fay,A., Chen,V.B., Block,J.N., Kapral,G.J., Wang,X., Murray,L.W., Arendall,W.B., Snoeyink,J., Richardson,J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–83.
26. Agari,Y., Sakamoto,K., Tamakoshi,M., Oshima,T., Kuramitsu,S. and Shinkai,A. (2010) Transcription profile of Thermus thermophilus CRISPR systems after phage infection. *J. Mol. Biol.*, **395**, 270–281.
27. Juranek,S., Eban,T., Altuvia,Y., Brown,M., Morozov,P., Tuschl,T. and Margalit,H. (2012) A genome-wide view of the expression and processing patterns of Thermus thermophilus HB8 CRISPR RNAs. *RNA*, **18**, 783–794.
28. Wang,R., Zheng,H., Preamplume,G., Shao,Y. and Li,H. (2012) The impact of CRISPR repeat sequence on structures of a Cas6 protein-RNA complex. *Protein Sci.*, **21**, 405–417.
29. Shao,Y. and Li,H. (2013) Recognition and cleavage of a nonstructured CRISPR RNA by its processing endoribonuclease Cas6. *Structure*, **21**, 385–393.
30. Reeks,J., Naismith,J.H. and White,M.F. (2013) CRISPR interference: a structural perspective. *Biochem. J.*, **453**, 155–166.
31. Lee,H.Y., Haurwitz,R.E., Apffel,A., Zhou,K., Smart,B., Wenger,C.D., Laderman,S., Bruhn,L. and Doudna,J.A. (2013) RNA-protein analysis using a conditional CRISPR nuclease. *Proc. Natl Acad. Sci. USA*, **110**, 5416–5421.
32. Makarova,K.S., Aravind,L., Wolf,Y.I. and Koonin,E.V. (2011) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct.*, **6**, 38.
33. Haft,D.H., Selengut,J., Mongodin,E.F. and Nelson,K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.*, **1**, e60.
34. Serra,M.J., Lyttle,M.H., Axenson,T.J., Schadt,C.A. and Turner,D.H. (1993) RNA hairpin loop stability depends on closing base pair. *Nucleic Acids Res.*, **21**, 3845–3849.
35. Makarova,K.S., Aravind,L., Grishin,N.V., Rogozin,I.B. and Koonin,E.V. (2002) A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.*, **30**, 482–496.
36. Wiedenheft,B., van Duijn,E., Bultema,J.B., Bultema,J., Waghmare,S.P., Waghmare,S., Zhou,K., Barendregt,A., Westphal,W., Heck,A.J.R. *et al.* (2011) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl Acad. Sci. USA*, **108**, 10092–10097.
37. Wiedenheft,B., Lander,G.C., Zhou,K., Jore,M.M., Brouns,S.J.J., van der Oost,J., Doudna,J.A. and Nogales,E. (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*, **477**, 486–489.
38. Hale,C.R., Zhao,P., Olson,S., Duff,M.O., Graveley,B.R., Wells,L., Terns,R.M. and Terns,M.P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, **139**, 945–956.
39. Lintner,N.G., Kerou,M., Brumfield,S.K., Graham,S., Liu,H., Naismith,J.H., Sdano,M., Peng,N., She,Q., Copie,V. *et al.* (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem.*, **286**, 21643–21656.
40. Haurwitz,R.E., Sternberg,S.H. and Doudna,J.A. (2012) Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *EMBO J.*, **31**, 2824–2832.
41. Grissa,I., Vergnaud,G. and Pourcel,C. (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, **8**, 172.
42. Velankar,S., Alhroub,Y., Best,C., Caboche,S., Conroy,M.J., Dana,J.M., Fernandez Montecelo,M.A., van Ginkel,G., Golovin,A., Gore,S.P. *et al.* (2012) PDBe: protein data bank in Europe. *Nucleic Acids Res.*, **40**, D445–D452.
43. Gouet,P., Courcelle,E., Stuart,D.I. and Métoz,F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.
44. Holm,L. and Sander,C. (1995) Dali: a network tool for protein structure comparison. *Trends Biochem Sci.*, **20**, 478–480.